

Andreas Degkwitz\*

# Die Tradition des Textes und die Herausforderung der Daten

DOI 10.1515/bfp-2017-0022

**Zusammenfassung:** Im Zuge der digitalen Transformation von Forschung, Lehre und Studium werden textorientierte Arbeitsformen zunehmend von datenbasierten Herangehensweisen abgelöst. Die sich damit verändernde Informationslogistik von Forschungsprozessen und Ergebnispräsentationen erfordert eine Neuausrichtung der bibliothekarischen Services. Denn die Dynamik datengetriebener, digitaler Prozesse bringt die Statik textgebundener, analoger Wissensarchitekturen ins Wanken. Mit den „Digital Humanities“ ist diese Entwicklung auch bei den Geistes- und Kulturwissenschaften angekommen. Digitale Informationsinfrastrukturen verändern bestehende Forschungs- und Publikationskulturen und führen zu neuen Formen der Produktion und Präsentation von Wissen. Der Beitrag zeigt die Veränderungen und Wechselwirkungen auf, die sich aus dem Paradigmenwandel von „Texten“ zu „Daten“ ergeben.<sup>1</sup>

**Schlüsselwörter:** Digitalisierung; Informationsinfrastrukturen; Datenbasiertes Wissen; Wissenschaftskommunikation; Digitisation; Information Infrastructures; Data Driven Knowledge; Scholarly Communication

## The Tradition of Texts and the Challenge of Data

**Abstract:** Text-oriented working procedures are turning to data-based approaches by the digital transition of education and research. New library services have to be established to meet the requirements of the changing information logistics of research processes and research results. Because the data driven dynamics of digital research is challen-

ging the analogue structures of text-based research. The digital humanities make clear, that these developments have reached the humanistic and cultural studies. Digital information infrastructures are changing existing research and publication cultures. New procedures of producing and presenting research results are rising up. The contribution outlines these changes and interdependencies as the result of the paradigmatic turn from texts to data.

**Keywords:** Digitisation; information infrastructures; data driven knowledge; scholarly communication

Forschung, Lehre und Studium sind seit Jahrhunderten geprägt von einer Kultur des Textes, die eine ebenso lange Tradition gedruckter, papiergebundener Inhalte repräsentiert. Die damit verbundene Logistik der Produktion, Verbreitung und Nutzung wissenschaftlicher Fachinformation sowie die Wertschöpfungsprozesse der Wissenschaftskommunikation („scholarly communication“) sind uns deshalb bestens vertraut. Die Organisationsstrukturen wissenschaftlicher Bibliotheken, die über ebenfalls lange Zeiträume ausschließlich an der Speicherung und Vermittlung analoger Medien orientiert waren, wurden und werden noch immer ganz wesentlich von dieser Informationslogistik bestimmt. Dazu gehören geschlossene und offene Magazine für gedruckte Bücher und Zeitschriften, das oft nur teil-automatisierbare und deshalb personalintensive Handling von Akquisitions-, Erschließungs- und Bereitstellungsprozessen sowie die durchaus auch über den Bibliotheksbereich hinaus bestehende Auffassung, dass Staats- und Universitätsbibliotheken gleichsam „Dienstleistungsmonopole“ im Kosmos der wissenschaftlichen Literatur- und Informationsversorgung sind. Doch diese „Welt“ befindet sich im Umbruch, seit die Medien digital, das WEB interaktiv und Texte zu Daten geworden sind. Zugleich wurde erfolgreich damit begonnen, neue Infrastrukturen für die Bereitstellung und Nutzung der mehr und mehr elektronisch verfügbaren Wissenschaftskommunikation zu entwickeln und aufzubauen. Der damit verbundene, tiefgreifende Transformationsprozess macht die großen Herausforderungen deutlich, die sich für Forschung, Lehre, Studium und für die Einrichtungen der akademischen Informationsinfrastrukturen daraus er-

<sup>1</sup> Dieser Beitrag ist eine gekürzte und in seinem Einleitungsteil und seiner abschließenden Zusammenfassung auf die Thematik des vorliegenden Bandes ausgerichtete Fassung meines Artikels: Transformation des Wissens und Wandel des Publizierens. In: *Struktur und Funktion wissenschaftlicher Publikation im World Wide Web: Wissenschaftsforschung Jahrbuch 2015*, hrsg. v. Klaus Fuchs-Kittowski, Heinrich Parthey und Walther Umstätter. Berlin: Wissenschaftlicher Verlag Berlin, 25–38.

\*Kontaktperson: Andreas Degkwitz,  
andreas.degkwitz@ub.hu-berlin.de

geben. Denn der Wandel der Medien, der sich im Übergang von statischen, textbasierten Wissenscontainern zu dynamischen, datenbasierten Wissensströmen vollzieht, betrifft organisatorische und technische Neuerungen und darüber hinaus die Kultur und das Verständnis der Produktion, der Verbreitung und der Aneignung von Wissen. Zugleich erweisen sich Infrastrukturentwicklung, „data driven research“ und digitale „scholarly communication“ als – sich wechselseitig beeinflussende – Treiber dieser Dynamik.

## 1 Digitalisierung der Informationsinfrastrukturen

Die Digitalisierung im Alltag unterschiedlicher Lebenswelten hat in Verbindung mit einer intensiven Nutzung des Internets zu einer massiven Überflutung der Verbrauchermärkte mit Endgeräten, mobilen Devices, Peripheriegeräten, Apps, Tools und einer Vielzahl von Software-Komponenten geführt. Die sich daraus ergebenden Innovationen liegen schwerpunktmäßig in der explosionsartigen Generierung, Verbreitung und Rezeption beliebiger Informationen und Inhalte sowie in der signifikant expandierenden, digitalen Kommunikation, die weltweit nahezu alle Ebenen des gesellschaftlichen Lebens durchdringt. In Forschung und Lehre hat sich mit der Digitalisierung ebenfalls ein Paradigmenwechsel vollzogen. Dazu gehören die Nutzung von Computer und Internet für die wissenschaftliche Arbeit sowie im Kontext der Bürokommunikation, die Digitalisierung der Wissenschaftskommunikation mit elektronischen Büchern und Zeitschriften unter Einschluss von retro-digitalisierten Materialien sowie die zahlreichen Applikationen – oder auch Geräte – zur Unterstützung von Arbeitsabläufen und Prozessen in Forschung, Lehre, Studium und Verwaltung. In diesen Zusammenhang gehören auch die IT-gestützten Verwaltungssysteme der Bibliotheken. War der Wissenschaftsbereich in den 80er- und 90er-Jahren des 20. Jahrhunderts zunächst Impulsgeber und Treiber für den Einsatz von IT-Anwendungen auch über Hochschulen und Forschungseinrichtungen hinaus, so hat sich diese Entwicklung nicht fortgesetzt, sondern wurde von der Marktentwicklung einer boomenden IT-Industrie eingeholt. Die Anbieter von mobilen Devices („smart technologies“), Kommunikationsplattformen („social networks“) und mächtigen Suchmaschinen einschließlich intuitiv nutzbarer Tools („app galleries“) haben die Entwicklungsführerschaft übernommen und bedienen mit hoher Dynamik einen weltweit präsenten Markt, an dem zahlreiche Content- und Service-

Provider unterschiedlicher Größenordnung partizipieren. Dabei ist entscheidend, dass die jeweils eingenommenen Marktpositionen Gewinn- und Umsatzpotenziale bieten, die den Firmen die wirtschaftliche Überlebensfähigkeit und den Marktanschluss trotz hoher Entwicklungsaufwände und kundenfreundlicher Preise sichern.

Die damit einhergehende Marktentwicklung mag verdeutlichen, wie sich die Ausrichtung der ursprünglich durchaus entwicklungsorientierten „Hochschul-IT“ von den Profilen der nun Markt beherrschenden Player unterscheidet. Anders gesagt: Die Weiterentwicklung der digitalen Transformation findet heute verstärkt außerhalb des Wissenschaftsbereichs statt, obwohl diese Entwicklung ursprünglich von Forschungseinrichtungen und Hochschulen ihren Ausgang nahm. Zugleich wird dadurch der Eindruck vermittelt, dass der vielfach diagnostizierte Innovationsdruck „von außen“ Forschung und Lehre nicht erreicht und insofern bisher nicht zu den notwendigen Veränderungen in der „Academia“ führte, an der sich Vorreiter von Innovation oder „early adopter“ eher in der Minderheit sehen und bisweilen sogar isoliert erscheinen. Etwas anders sieht dies bei zentralen Serviceeinrichtungen wie Bibliotheken, Medien- und Rechenzentren aus, die sich als Informations- oder IT-Dienstleister in der Situation befinden, die Herausforderungen der Digitalisierung aufzugreifen und sie mithilfe der Entwicklung entsprechender Dienstleistungen umzusetzen. Dabei wird allerdings immer wieder deutlich, dass für die Akzeptanz innovativer Informations- und Mediendienste viel Überzeugungsarbeit erforderlich ist.

Ganz unbestritten ist, dass der Wissenschaftsbereich andere Anforderungen an den Einsatz innovativer Technologie hat als die jeweils aktuelle und sich permanent wandelnde „Alltags-IT“, die schon aus fachlichen Gründen nicht unmittelbar auf den Wissenschaftsbereich übertragbar ist. Zugleich sind die Einrichtungen der Informationsinfrastruktur an Hochschulen aus finanziellen oder kapazitären Gründen oft nur eingeschränkt in der Lage, die Weiterentwicklung ihrer Serviceportfolios an die Entwicklungsdynamik marktgetriebener Innovationen anzupassen. Dabei sind über fehlende Ressourcen hinaus insbesondere rechtliche Rahmenbedingungen zu nennen, wie beispielsweise die Einhaltung datenschutzrechtlicher Regelungen oder die Gewährleistung von Datensicherheit. Diese Rahmenbedingungen führen oft dazu, dass hochschul- bzw. wissenschaftseigene Angebote unflexibel und zu wenig komfortabel erscheinen, um wirklich „gerne“ genutzt zu werden. Verglichen mit den Angeboten auf dem Verbrauchermarkt kommen schließlich viele Entwicklungen wissenschaftseigener Dienste und Infrastrukturen zu spät. Dies wiederum führt dazu, dass die im Regelfall ein-

fach nutzbaren, oft sehr preiswerten Angebote großer IT-Provider, wie insbesondere Kollaborationstools, Mail, Online-Speicher und „social networks“, doch in den Wissenschaftsbereich Eingang finden, ohne dass die im Grundsatz erforderlichen Voraussetzungen des Datenschutzes, der Datensicherheit und der Datenhoheit gegeben sind. In Anbetracht dessen ist und bleibt der Aufbau wissenschaftsspezifischer Infrastrukturen auch künftig ein dringendes Desiderat.

Vor diesem Hintergrund hat es in den vergangenen 20 Jahren weltweit auf internationaler und nationaler Ebene sowie innerhalb und außerhalb des Wissenschaftsbereichs eine Vielzahl an Empfehlungen, Förderinitiativen und Projekten gegeben, die die Herausforderungen der digitalen Transformation für Forschung, Lehre, Studium und Verwaltung aufgegriffen und in unterschiedlichem Umfang zur Realisierung gebracht haben. Die Stichwörter dieser Entwicklung sind für die Forschung „e-research“ und „e-science“, für Lehre und Studium „e-learning“ und „e-teaching“ und für den Bereich der Verwaltung „e-administration“ und „e-governance“. Dass die Vorsilbe „e-“ eingeführt wurde, um die „digitale Hochschule“ von der „analogen Hochschule“ zu unterscheiden und die herkömmlichen Verfahren von den neuen, technologiegetriebenen Verfahren im Sinne von „enhanced“ abzusetzen, soll hier ebenso wenig vertieft werden wie der immer wieder neue Versuch, die Veränderungspotenziale der Informationstechnologie für die Einführung unternehmerischer Strukturen im Sinne von „business models“ in Forschung, Lehre und Studium einzusetzen. Was vor allem mit der Hochschule als unternehmerischem Geschäftsmodell adressiert wird, ist ein Kulturwandel, der im Zuge der Digitalisierung des akademischen „Kerngeschäfts“ möglich erscheint und dabei zu einem veränderten, neuen Selbstverständnis der „academia“ führen kann. Im vorliegenden Zusammenhang soll dagegen der mit der digitalen Transformation einhergehende Wandel des Wissensverständnisses betrachtet werden, der im direkten Zusammenhang mit der Weiterentwicklung der Informationsinfrastrukturen steht. Denn bei den Aktivitäten und Initiativen, die unter dem Label „enhanced education and enhanced research“ anvisiert und umgesetzt werden, ist eine unmittelbare Verbindung mit der jeweils erforderlichen Informationsinfrastruktur gegeben, die zwingende Voraussetzung für die digitale Transformation von Forschung, Lehre und Studium ist.

Vor diesem Hintergrund bestehen digitale Informationsinfrastrukturen zum einen im Sinne einer eher informationszentrierten Unterstützung des wissenschaftlichen Arbeitens. Dabei geht es um Verfügbarkeit und Verbreitung von Inhalten, die selbst in digitaler Form vorliegen

oder als digitale Objekte generiert werden. Die dafür notwendige Unterstützung wird von Bibliotheken angeboten, die E-Books und E-Journals lizenzieren, Recherche- und Beratungsservices anbieten, Forschungsergebnisse nach den Prinzipien des Open Access publizieren, analog verfügbare Materialien im Rahmen der Möglichkeiten der Urheberrechtsgesetzgebung (retro-)digitalisieren und die sich um die langfristige Verfügbarkeit bzw. Archivierung digitaler Inhalte (Bilder, Daten, Texte etc.) bemühen. Zum anderen – und darüber hinaus gehend – sind künftig verstärkt von Bibliotheken anzubietende Infrastrukturdienste zu nennen, die kollaborative Arbeitsformen und die Aufbereitung und Verarbeitung von Inhalten durch Annotieren, Kontextualisierung, Referenzierung und Strukturierung sowie Data- und Textmining oder andere Retrievalformen unterstützen. Dabei geht es im Unterschied zu den zuvor genannten Unterstützungsmaßnahmen um Dienste zur informationstechnischen Unterstützung, die sich unmittelbar auf das wissenschaftliche Arbeiten beziehen und damit fester Bestandteil von Forschungsprozessen sind. Dabei wird der epistemische Zusammenhang zwischen Infrastruktur und Wissensgenerierung deutlich, der im Grundsatz nicht neu ist, aber in digitaler Form eine neue Qualität enthält. Denn maschinelle Arbeits- und Auswertungsverfahren ersetzen analoge oder manuelle Prozesse und enthalten ein ungleich größeres Potenzial an Effektivität und Effizienz, als dies mit herkömmlichen Verfahren der Verarbeitung und Evaluation möglich und vorstellbar ist.

## 2 Wissen im Wandel

Wenn IT-gestützte Informationsinfrastrukturen eng mit der Produktion, der Repräsentation und der Rezeption von Wissen und Wissensgütern in Zusammenhang stehen, werden die Prozesse der wissenschaftlichen Arbeit verstärkt automatisiert und damit maschinellen Verfahren unterworfen. Solche Verfahren sind in der Geschichte des Wissens nicht ungewöhnlich und geschehen mit der Digitalisierung gewiss nicht zum ersten Mal. Denn schon immer haben Menschen sich darum bemüht, große Herausforderungen an Technik auszulagern. Ein sehr gutes Beispiel dafür ist das Automobil, an das vor allem körperliche Bewegung von Stadt A nach Stadt B ausgelagert wird. Seit vielen Jahrhunderten – um nicht zu sagen, seit einigen Jahrtausenden – lagern wir unser Gedächtnis zunächst an die Technik der Schrift, dann an beschriftete Trägermedien (Papyrus, Pergament, Papier) und nun an Maschinen, Computer und Netze aus. Indem wir Verarbeitung („processing“), Verbreitung („distribution“) und Speicher („me-

mory“) auf Server und Netze auslagern können, sind wir in der Situation, in einem Umfang auszulagern, der uns bisher nicht zur Verfügung stand. Wir konnten noch nie so viel an Gedächtnis, Verarbeitung und Verbreitung auslagern, wie es uns heute möglich ist.<sup>2</sup> Aber wir laufen zugleich Gefahr, durch Auslagerung auch zu verlieren, so dass sich die Frage stellt: Was machen wir mit diesen neuen Maschinen und was machen diese Maschinen mit uns? Mit dieser Frage wird deutlich, dass die digitale Transformation mit Entwicklungen der künstlichen Intelligenz („artificial intelligence“) in Zusammenhang steht, in der mit informationstechnischen Ansätzen versucht wird, kreative Intelligenz zu modellieren und mithilfe technischer Verfahren zu automatisieren. Mag ein in diesem Sinne vollständiger Ansatz die Möglichkeiten der künstlichen Intelligenz überschreiten, so zählen einzelne Komponenten wie Information Retrieval, Suchmaschinen, Sprach- und Zeichenerkennung, Text- und Datamining, maschinelles Übersetzen oder wissensbasierte Systeme zu ihren Anwendungsfeldern, die als Teilgebiete der künstlichen Intelligenz beforscht wurden und werden.<sup>3</sup>

Unter epistemischen Gesichtspunkten wird im Zusammenhang mit der Digitalisierung wissenschaftlicher Arbeitsformen von einer „neuen“ Qualität des Wissens gesprochen, indem mit den Möglichkeiten digitaler Verarbeitung und Vernetzung von Daten und Texten – nicht zuletzt auch für Fragen und Themen, die bereits als erforscht gelten – neue Erkenntnisziele identifiziert und damit neue Forschungsergebnisse gewonnen werden können, die aufgrund der Vollständigkeit der jeweils zugrunde liegenden Datenbasis und der darauf aufsetzenden Werkzeuge mit analogen Methoden nicht hätten erreicht werden können. Erkenntnisziele und Wissensergebnisse stellen sich in Geistes- und Naturwissenschaften ganz unterschiedlich dar. In den Naturwissenschaften, die von beobachtenden, experimentierenden oder simulationsbezogenen Forschungsformen<sup>4</sup> stark geprägt sind, bilden Daten in nahezu allen Formen die Grundlage wissenschaftlicher Arbeit. Von daher werden seit langem Daten und Datenbestände mithilfe der „computational science“ analysiert, evaluiert, simuliert und visualisiert, um anhand entsprechender Modelle, die an beobachteten Phänomenen orientiert sind, Kausalitäten und Kontexte zu erklären und um

Gesetzmäßigkeiten abzuleiten oder Entwicklungen zu prognostizieren.<sup>5</sup> Von daher wird in den Naturwissenschaften vorrangig mit Ansätzen der Statistik und Wahrscheinlichkeitsverteilung gearbeitet. Im Unterschied zu naturwissenschaftlichen Forschungen steht für die Geistes-, Kultur- und Sozialwissenschaften primär das Verstehen von menschlichem Denken, Handeln, Schaffen und Wirken im Mittelpunkt. Vor diesem Hintergrund stellt sich die Frage, ob diese auf Daten beruhenden Ansätze auch auf hermeneutisch-interpretierende Forschungsformen übertragbar sind und sich in den „Digital Humanities“ mit diesen Methoden „Verstehenszusammenhänge“ sinnvoll erforschen lassen. Denn wenn datenbasierte Arbeitsverfahren in den „humanities“ und in den „social sciences“ eingesetzt werden, steht zu erwarten, dass die hermeneutisch-interpretierenden Forschungsformen, die für diese Wissensgebiete einen hohen Stellenwert haben, absehbar durch empirische Methoden und quantitative Verfahren ergänzt und möglicherweise sogar durch diese ersetzt werden.<sup>6</sup>

Anknüpfend an die herkömmlichen, analogen Verfahren der hermeneutisch-interpretierenden Forschungsformen werden auch in den „Digital Humanities“ in einem umfassenden Sinne „Artefakte“ wie insbesondere Bilder und Texte zunächst gesammelt und ausgewertet (analysiert). Im digitalen Kontext stellen die technische Aufbereitung sowie die XML-Strukturierung der „Artefakte“ für maschinelle Analyse und Weiterverarbeitung eine wesentliche Voraussetzung dar. Gegenüber analogen Arbeitsweisen sind Aufbereitung und Strukturierung im digitalen Zusammenhang ein zusätzlicher und neuer Teilprozess. Im nächsten Schritt werden die Beschreibungen und Analyseergebnisse des Bild- und Textmaterials referenziert und semantisch in Beziehung gesetzt (kontextualisiert). Schließlich werden die Ergebnisse der Analyse-, Evaluations- und Kontextualisierungsverfahren aufbereitet und veranschaulicht (visualisiert), um anhand der für das Forschungsvorhaben herangezogenen Artefakte die erkenntnisleitenden Fragestellungen zu menschlichem Verhalten und schöpferischem Wirken zu beantworten und in ihrer Kausalität zu verstehen. Datenbasierte Analysen, die quantitative Auswertungen, semantische Kontextualisierung und die Visualisierung von Ergebnissen ermöglichen, können dabei ganz wesentlich zu neuen Erkenntnissen beitragen wie auch zu neuen Forschungsfragen führen.<sup>7</sup> Die Herausforderungen, die sich dabei stellen, sind darin

<sup>2</sup> Vgl. Degkwitz (2012) und Zimmerli (2000).

<sup>3</sup> Rehm und Uszkoreit (2012).

<sup>4</sup> Zu den in den Geistes- und Naturwissenschaften zu beobachtenden, experimentierenden, simulationsbezogenen, hermeneutisch-interpretierenden, begrifflich-theoretischen und gestaltenden Forschungsformen vgl. die Empfehlungen vom Wissenschaftsrat (2012) 35 ff.

<sup>5</sup> Siehe Zimmerli (2008) 22: „Galt bislang Georg Simmels Bestimmung, gebildet sei, wer wisse, wo er finde, was er nicht weiß, gilt nun: Gebildet ist, wer weiß, wie er machen kann, was er nicht weiß.“

<sup>6</sup> Siehe Blanke und Horstmann (2013) 55 ff.

<sup>7</sup> Vgl. Görtz (2011).



zu sehen, dass die „Quelldaten“ der Geistes-, Kultur- und Sozialwissenschaften überwiegend nicht digital vorliegen, sondern zunächst zu digitalen Daten transformiert werden müssen, um in der weiteren Folge als Datenmaterial verarbeitet werden zu können. Dies gilt im Übrigen auch für digitale „Quelldaten“, die nicht in verarbeitungsfähigen Formaten zur Verfügung stehen. Mit anderen Worten geht es darum, dass das „Implizite“ der zunächst nicht inter-operablen bzw. nicht verarbeitbaren „Quelldaten“ geöffnet und „expliziert“ wird. Als Wissenschaften des Diskurses und Verstehens sind den Geisteswissenschaften damit eine Verbreiterung der bisher zugänglichen Datengrundlage und „eine Explizierung des in unseren Daten- und Wissensspeichern inhärenten Wissens verfügbar“, was – „verbunden mit einer gewissen Annäherung an naturwissenschaftliche Zugänge“ – für ihre Forschungsfragen gewinnbringend ist, ohne dabei etwas aufgeben zu müssen. Von daher verbinden sich mit dem Zugriff auf bisher nicht zugängliche Quellen oder einfach nur „mehr“ Quellen für die Geisteswissenschaften neue und eben „mehr“ Erkenntnisse.<sup>8</sup>

Zugleich stellt sich die Frage der Leistungsfähigkeit und der Transparenz von Algorithmen und Werkzeugen, die zur Verarbeitung der dann digital transformierten Daten eingesetzt werden.<sup>9</sup> Sind maschinelle Verfahren zur Analyse, Auswertung und Kontextualisierung für interpretierende Forschungsformen geeignet und angebracht? Ist uns bewusst, was auf den Maschinen passiert, auf die wir unser „Gedächtnis“ und die Verarbeitung unserer „Erinnerung“ auslagern? Möglicherweise ist dabei nicht so sehr der Einsatz maschineller Verfahren kritisch, sondern vielmehr ein „blindes“ und weitgehend kenntnisfreies Vertrauen auf eine hohe Leistungsfähigkeit und Objektivität der eingesetzten Verfahren, mit denen Daten maschinell generiert oder maschinell generierte Daten genutzt werden, die dann die Grundlage für Analysen und Auswertungen sind. Denn auf diese Weise wird eine Realität geschaffen, die keine konkrete, physische Realität, keine spekulativ gebrochene Realität, sondern eine virtuelle Realität ist – eine Realität, die maschinell simuliert ist und von daher in ihrem Wesen und Wirken der physischen Realität gleichgesetzt wird und dabei auf Daten und Datenbeständen beruht, die die Grundlage für die Verarbeitung und Aggregation empirischen Wissens sind. Anders gesagt: Wir wissen, indem wir generieren, was wir nicht wissen und kommen auf diese Weise zu Ergebnissen, die wir als Empirie verstehen. Dass in dieser virtuellen Umgebung auch die Gefahr besteht,

Wirklichkeitsbezüge nur eingeschränkt herzustellen oder gar zu verlieren, ist der Preis der Auslagerung unseres „Gedächtnisses“ an Maschinen.

Schließlich bieten uns solche virtuellen Arbeitsumformen gewaltige Mehrwerte, indem sie eine Mehrfachverarbeitung von Datenbeständen erlauben, die uns – wie schon gesagt – neue Forschungsfragen und Wissensgebiete eröffnen. Alle Daten, die in dieser Weise quantitativ verarbeitet oder semantisch kontextualisiert werden, können zu neuen Erkenntnissen oder zu neuen Fragestellungen führen, die weitere Verarbeitungsprozesse veranlassen und neue Forschungsprojekte – auch für hermeneutisch-interpretierende Fachgebiete – initiieren. Für die Produktion, Transparenz und Verbreitung von Wissen und Wissensgütern ist das ein enormer Gewinn.<sup>10</sup> Die jeweiligen Eigenheiten und paradigmatischen Unterschiede der Geistes- und Naturwissenschaften werden dabei weder ausgeräumt noch gleichgeschaltet, sondern bleiben bestehen. Doch bieten sich neue Möglichkeiten, Methoden und Verfahren wechselseitig zu übernehmen, weil nun auch in den Geisteswissenschaften Daten und Datenbestände die Grundlage der wissenschaftlichen Arbeit sind.

### 3 Forschungsprozess und Publikation

Oft wird der Wandel von analogen (gedruckten) zu digitalen (elektronischen) Medien als unmittelbarer Auslöser für die digitale, vernetzte Transformation von Informationen und Kommunikationsprozessen betrachtet. Zugleich unterliegt der Medienwandel dem Einfluss von IT-gestützten Arbeitsverfahren und Workflows, die entweder aus Gründen der Optimierung bestehender Prozesse oder für die Herstellung neuer Produkte eingesetzt werden, die sich mit herkömmlichen Verfahren nicht produzieren lassen. Für das wissenschaftliche Publizieren steht die Digitalisierung der Medien in direktem Zusammenhang mit dem Forschungsprozess, dessen Verlauf immer wieder zu Ergebnispräsentationen und damit zu Publikationen führt, die zugleich als mediale Präsentationen zu sehen sind. Denn Wissen und Medien hängen eng miteinander zusammen und beeinflussen sich wechselseitig: Ohne mediale Transformation kann Wissen weder vermittelt noch verbreitet werden. Medien repräsentieren unser Wissensverständnis. Wissensparadigmen prägen die Gestaltung der medialen Präsentation. Die mediale Gestaltung von Wissen beruht

<sup>8</sup> Siehe Rapp (2013) 347.

<sup>9</sup> Siehe Enke und Fritzsche (2013) 186.

<sup>10</sup> Siehe Blanke und Horstmann (2013) 63f. und Rapp (2013) 349.

deshalb nicht nur auf Technologie, sie ist auch in den unterschiedlichen, disziplinbezogenen Wissenskulturen begründet.

Über viele Jahrhunderte wurde im europäischen Kontext Wissen in den unterschiedlichen Formaten gedruckter Texte zusammengeführt und veröffentlicht. Ausschließlich textbasiertes Wissen und seine gedruckten Präsentationen verlieren jedoch an Bedeutung. Stattdessen entstehen verstärkt datengestützte Publikationsformen, die multimedial angereichert sind und sich in beliebige Kollaborationskontexte einbringen lassen oder auf kontextualisierten Daten unterschiedlicher Provenienz basieren. In diesem Zusammenhang gehören beispielsweise Druckbestände, die im OCR-Verfahren als Volltexte digitalisiert sind und für ihre weitere Verarbeitung nach XML strukturiert werden. Darüber hinaus bietet die Nutzung von Softwarekomponenten aus App-Galleries oder App-Stores die Möglichkeit, datenbasiertes Wissen unter verschiedenen Aspekten und Fragestellungen zu verarbeiten. Unterschiedliche Informationsmaterialien können in einem digitalen Dokument zusammen geführt werden und damit Bestandteile einer weiter bearbeitbaren Publikation bilden oder sie werden als differenziert verlinkte Datenbestände veröffentlicht. Die neuen, digitalen Medienformate übersteigen damit das Potenzial gedruckter Formate signifikant. Dies hat Auswirkungen auf Vermittlung und Verbreitung künftigen Wissens. Hinzu kommt das außerordentlich hohe Potenzial an Kommunikationsmöglichkeiten des Internets, das als Informationsreservoir weltweit just in time zur Verfügung steht. Die „single channel communication“ gedruckter Publikationen wird zunehmend durch die „networked communication“ digitaler Medien ersetzt. Der tiefgreifende Wandel hat das Potenzial, Dokumente und Texte nun zu multimedialen Präsentationen zu erweitern und den linearen „step by step process“ der gedruckten Veröffentlichungspraxis durch die „all to all communication“ des Internets als primärer Verbreitungsform für wissenschaftliche Publikationen zu ersetzen.

Dies lenkt den Blick auf den Forschungsprozess, der auch in den Geistes- und Kulturwissenschaften als genuin IT-gestützter Prozess nicht mehr ausschließlich textbasiert, sondern zunehmend auf der Grundlage von Daten erfolgt. Grundlegende Phasen des Forschungsprozesses aller Fachdisziplinen und Forschungsformen sind vor dem Hintergrund der jeweiligen Forschungsfragen (1) die Generierung oder Sammlung von Materialien, (2) deren Aufbereitung, Strukturierung und Aggregation, (3) Analyse und Evaluation der Materialien und (4) die Ableitung und Gewinnung von Ergebnissen und deren Publikation. Wenn diesen Phasen des Forschungsprozesses Daten und Datenbestände als Materialien zugrunde gelegt werden,

ergeben sich daraus Arbeitsvorgänge und Verfahren, die den Forschungsprozess bis zur Publikation unterstützen. Dabei sollen und werden verarbeitete Daten zugleich auch Bestandteil der Publikation sein und die Veröffentlichung der Ergebnispräsentation prägen. Denn für die Generierung oder Sammlung von Materialien werden Daten entweder produziert oder gewonnen oder mithilfe von Suchwerkzeugen identifiziert. Zur Aufbereitung und Aggregation von Materialien müssen Daten beschrieben, referenziert und strukturiert werden. Für Analyse und Evaluation werden Retrievalverfahren, wie beispielsweise Data- und Textmining, eingesetzt. Zur Ergebnisaufbereitung und Veröffentlichung werden über textuelle Ergebnispräsentationen hinaus Visualisierungen und multimediale Anreicherungen genutzt. Hinzu kommen die vielfältigen Möglichkeiten, Publikationen als „enhanced publications“, multimedial anzureichern, zu kommentieren, zu verlinken und weiter zu verarbeiten.<sup>11</sup> Da es bei datengetriebenen Forschungsprozessen und datenbasierten Ergebnisveröffentlichungen um die Speicherung und Verarbeitung großer Datenmengen („big data“) geht, sind hierfür virtualisierte und hochgradig skalierbare Systemumgebungen („clouds“) unerlässlich.

Damit gehen Veränderungen der Publikationskette einher, für die weiterhin das Peer-Reviewing eine zentrale Rolle spielt, deren Prozesse jedoch nicht mehr aufeinander aufbauen, sondern sehr viel stärker als bisher in einem Vernetzungskontext stehen. Auch wird und kann die am Ende stehende Publikation auf verschiedenen Wegen der Wissenschaftskommunikation distribuiert werden, was insbesondere in Verbindung mit dem Open-Access-Publizieren zu einer erhöhten Wahrnehmbarkeit von Forschungsergebnissen führt. Diese deutlich gesteigerten Möglichkeiten der Verbreitung und Vermittlung von Inhalten können und sollten nun noch dadurch erweitert werden, dass mit „enhanced publications“ Inhalte zur Verfügung stehen, die maschinell verarbeitet und somit nachgenutzt werden können. Dies gehört entweder zu den künftigen Aufgaben von Verlagen, die die zur Veröffentlichung vorgesehenen Inhalte entsprechend aufbereiten sollten, oder zum Servicespektrum von Bibliotheken, die die Aufbereitung von Veröffentlichungen nach ihrem Erscheinen vollziehen. Entscheidende Voraussetzungen dafür sind rechtliche Rahmenbedingungen, die die Aufbereitung und Verarbeitung publizierter Forschungsergebnisse erlauben. Doch an dieser Stelle stehen Bibliotheken und Verlage noch ganz am Anfang. Mit der Verarbeitbarkeit von Inhalten sind insbesondere Bibliotheken in der Situation,

<sup>11</sup> Vgl. Degkwitz (2014).

(1) entsprechende Werkzeuge zur Verfügung zu stellen, (2) für Metadatenbeschreibung, Textstrukturierung und Referenzierung von Inhalten und Objekten geeignete Dienste bereitzustellen und (3) im Rahmen ihres Sammlungs- und Versorgungsauftrages die langfristige Verfügbarkeit zu gewährleisten. Das sind schon jetzt und auf jeden Fall künftig zentrale Erwartungen an digitale Bibliotheken und ihre Sammlungen.

## 4 Zusammenfassung

Als Serviceeinrichtungen für die Bereitstellung und Nutzung publizierter Forschungsergebnisse stehen wissenschaftliche Bibliotheken am Ende der Wertschöpfungskette der Wissenschaftskommunikation und zugleich an deren Anfang, wenn es im Kontext von Forschungsvorhaben um die Recherchen zum Stand der Forschung geht. Im Zuge der digitalen Transformation verlagert sich das Aufgabenspektrum von Bibliotheken zunehmend auf den Forschungsprozess, für den verstärkt unterstützende Infrastrukturdienste erwartet werden und erforderlich sind. Dabei geht es oftmals um einen Wandel der Forschungs- und Publikationskulturen, der in den stärker auf die Verarbeitung von Daten ausgerichteten Naturwissenschaften weiter vorangeschritten ist als in den eher an Texten orientierten, bibliotheksaffinen Fachgebieten der Geistes-, Kultur- und Sozialwissenschaften. Mit diesem Kulturwandel verlieren traditionelle Arbeitsabläufe und Vorgehensweisen gegenüber neuen Methoden und Verfahren ihren bisherigen Stellenwert – dies betrifft Forschung, Lehre und Studium in derselben Weise wie die Einrichtungen der akademischen Informationsinfrastruktur. Die Dynamik datengetriebener, digitaler Prozesse bringt die Statik textgebundener, analoger Wissensarchitekturen ins Wanken, indem die in Druckwerken materialisierten Wissensbausteine Schritt für Schritt durch virtuelle, vernetzte „knowledge chunks“ ersetzt werden. Hochschulen, Forschungseinrichtungen und ihre Infrastruktureinrichtungen sind deshalb in der Situation, die damit verbundenen Herausforderungen aufzugreifen, die ihre Organisation, ihre Wertschöpfungsprozesse und ihre Wissenschaftskommunikation signifikant verändern werden – dies kann nur in eng erfolgreicher Abstimmung und Kooperation der daran beteiligten Akteure erfolgreich in Angriff genommen und nachhaltig realisiert werden.

## Literaturverzeichnis

- Blanke, Tobias; Horstmann, Axel (2013): Bericht über das Symposium zum zehnjährigen Bestehen der Abteilung Forschung und Entwicklung. In: *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft*, hg. v. Heike Neuroth, Norbert Lossau und Andrea Rapp, 39–70. Glückstadt: Verlag Werner Hülsbusch.
- Degkwitz, Andreas (2012): Texte, Daten, Bilder – Wissen! In: *BIBLIOTHEK – Forschung und Praxis*, 36 (3), 215–19.
- Degkwitz, Andreas (2014): Digitale Sammlungen – Vision eines Neubeginns. In: *BIBLIOTHEK – Forschung und Praxis*, 38 (3), 411–16.
- Enke, Harry; Fritsch, Bernadette (2013): Erfahrungen aus dem Projekt „WissGrid – Grid für die Wissenschaft“. In: *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft*, hg. v. Heike Neuroth, Norbert Lossau und Andrea Rapp, 179–93. Glückstadt: Verlag Werner Hülsbusch.
- Görtz, Günther (2011): KI und Geisteswissenschaften. In: *KI – Künstliche Intelligenz. Organ des Fachbereichs „Künstliche Intelligenz“ der Gesellschaft für Informatik e. V.*, 25 (4), 313–15.
- Rapp, Andrea (2013): Aus Sicht der Geisteswissenschaften: Die wissenschaftliche Bibliothek als Schnittstelle zwischen digitaler Infrastruktur und geisteswissenschaftlicher Forschung. In: *Evolution der Informationsinfrastruktur: Kooperation zwischen Bibliothek und Wissenschaft*, hg. v. Heike Neuroth, Norbert Lossau und Andrea Rapp, 345–54. Glückstadt: Verlag Werner Hülsbusch.
- Rehm, Georg; Uszkoreit, Hans (Hrsg.) (2012): *The German Language in the Digital Age*. Berlin, Heidelberg: Springer (White Paper Series).
- Wissenschaftsrat (2012): Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. Wissenschaftsrat. Verfügbar unter <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>.
- Zimmerli, Walther (2000): Vom Unterschied, der einen Unterschied macht. Information, Netzwerkdenken und Mensch-Maschine-Tandem. In: *Die Zukunft des Wissens. XVIII. Deutscher Kongress für Philosophie (1999)*, hg. v. Jürgen Mittelstrass, 120–22. Berlin: Akademie Verlag.
- Zimmerli, Walther (2008): Wissenskulturen des 18. und 21. Jahrhunderts. In: *Kulturen des Wissens im 18. Jahrhundert*, hg. v. Johannes Schneider, 1–22. Berlin: De Gruyter.



**Andreas Degkwitz**

Universitätsbibliothek der  
Humboldt-Universität zu Berlin  
Unter den Linden 6  
D-10099 Berlin  
[andreas.degkwitz@ub.hu-berlin.de](mailto:andreas.degkwitz@ub.hu-berlin.de)